CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 629 – 633
Procedia Engineering
www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

HDP-HMM-SCFG: A Novel Model for Trajectory Representation and Classification

Weiguang Xu*, Yafei Zhang, Jianjiang Lu, Jiabao Wang

Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China

Abstract

In this paper, we propose a novel model, HDP-HMM-SCFG, for representing and classifying trajectories. Trajectories are represented by stochastic grammar, where trajectory segments are considered as observations emitted by the grammar terminals, which are attached with HMMs. In order to learn the parameters of SCFG, we employ hierarchical Dirichlet Processes (HDP) as the nonparameter prior of the distribution of the parameters, and obtain the model of HDP-HMM-SCFG. Then, we propose a 3-level CRP based Gibbs sampling inference algorithm to acquire the SCFG parameters. In the training phase of classification, SCFGs for different classes are learned respectively by inferring on the training sets with different labels independently. Then test trajectory is parsed with a bottom-up parsing algorithm, and the probability for each SCFG to generate it is calculated. The label of the class with the maximum likelihood to generate the test trajectory is assigned to the test trajectory. Experiment on ASL dataset is carried on to validate our approach.

© 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: HDP-HMM-SCFG; trajectory; representation; classification.

1. Introduction

Trajectory analysis is a basic task in the many domains, such as visual surveillance automation, intelligent human-computer interaction, and recognizing behavior of moving objects in radar fence. Raw trajectories are usually denoted by a sequence of points, $T=\{t_1,t_2,\ldots,t_n\}$, each of which is a 2D coordinate,

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: miipl606@163.com.

1877-7058 © 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2011.08.117

or sometimes higher dimensional coordinate. The raw trajectories are generally noisy, because of the inaccuracy of the sensors, and the natural instability of nonrigid objects. Besides the noise, the complex temporal relationship and hierarchical structure of different scale sub-trajectories is another obstacle. Therefore, trajectory analysis is a challenging problem and has attracted a lot of attentions in recent years. In this paper, a novel model of HDP-HMM-SCFG is proposed for representing and classifying trajectories. The model is a synthesis of nonparametric Bayesian model and grammatical model.

The rest of the paper is structured as follows: in section 2, we discuss the related works of trajectory analysis. In section 3, we detailed our model of HDP-HMM-SCFG in three main aspects: representation, inference, and classification. In section 4, we introduce the experiment on ASL dataset. At last, a conclusion is drawn in section 5, and some future works are also discussed.

2. Related Work

Most of the previous works analysed trajectories in three steps: preprocessing, clustering, and modeling. The step of preprocessing is to normalize and reduce the dimensionality of the raw trajectories. The lengths of trajectories are not equal at most scenes. Zero padding [1] and track extension [2, 3] are the most popular and simple way to normalize them. Due to the high dimensionality of raw trajectory, Vector quantization reduction [4, 5], HMM [6] and PCA [6] are usually employed to reduce the dimensionality. The step of clustering learns the different patterns from the unlabeled trajectories by employing various distance measures (Euclidean, DTW [7], LCSS [8, 9], EMD) and clustering algorithms. Clustering algorithms have been reviewed by Berkhin and Warren [10, 11]. The step of modeling models the discovered clusters with their centroids and envelopes. The centroid denotes the prototype of the cluster, and the envelope depicts the extent of the cluster. GMMs and HMMs [6] are the most common way that models trajectory clusters. However, GMMs are lack of temporal information and HMMs lacks of hierarchical information.

Besides the traditional approaches mentioned above, grammar based models have come forth to the domain of computer vision recently due to its success in speech recognition [12]. However, the traditional grammar based models needs experts to predefine the production rules and other parameters. Liang et al. [13] combined the probabilistic grammars and nonparametric model of Hierarchical Dirichlet Process to analyze natural language in an unsupervised manner. Their work gives a solution of learning a probabilistic grammar from training set of sequences.

3. Representation, Inference and Classification

In our approach, a model of HMM-SCFG, combining HMM with stochastic grammar, is proposed to represent trajectory cluster. And then, nonparametric model HDP is adopted to obtain the model of HDP-HMM-SCFG. A Gibbs sampling inference algorithm based on 3-level Chinese Restaurant Process (CRP) is proposed to estimate the parameters. Maximum Likelihood strategy is used to classify a test trajectory into a category at last.

3.1. Representation with HDP-HMM-SCFG

The HMM-SCFG is formally represented by a five-tuple: $G=(N, T, A, P, S)$, where the N is the set of finite nonterminals; T is the set of finite terminals, each of which combined with a HMM; A is the set of finite trajectory segments, which is a continuous infinite space; P is the set of finite production rules, each of which combined with a probability, and $S\in N$ is the root nonterminal. The difference with the normal SCFG is that the ultimate nodes during derivation are not the terminal symbols but the trajectory segments, which are emitted by terminal symbols. The parameters need to be learned include the nonterminals, the terminals and their HMMs, the production rules and their probabilities. The problem is the number of nonterminals, terminals, HMMs components, and rules are all unknown. To solve this, we employ hierarchical Dirichlet Processes (HDP) as the nonparameter prior of the distribution of the parameters, and then lead to the model of HDP-HMM-SCFG. The number of nonterminals, terminals, production rules, and HMMs components are all infinite in the model. All their prior distributions are specified as follows.

$\pi_0^{non\_tmn} \sim GEM(\alpha^{non\_tmn})$	
$N \sim multi(\pi_0^{non\_tmn})$	//prior distribution of all the nonterminals
$\pi_0^{son} \sim GEM(\alpha^{son})$	
$\theta_k^P : DP(\alpha^P, \pi_0^{son} \cdot (\pi_0^{son})^T), k = 1,2,...,\infty$	//prior distribution of all the production rules
$\pi_0^{tmn} \sim GEM(\alpha^{tmn})$	
$\theta_k^{Tran} : DP(\alpha^{Tran}, \pi_0^{tmn}), k = 1,2,...,\infty$	//prior distribution of all the HMMs transitions
$\theta_k^E : W^{-1}(M^E), k = 1,2,...,\infty$	//prior distribution of all the HMMs components

Where the *GEM* is the stick breaking distribution [14]; *multi* is the multinomial distribution; *DP* is the Dirichlet Process, W^1 is the Inverse-Wishart distribution, which is the conjugate prior for the Gaussian distribution. α^{non_tmn}, α^{son}, α^{tmn}, α^{Tran} and M^E are predefined super parameters. $\pi_0^{non\_tmn}$, θ_k^P, θ_k^{Tran} and θ_k^E are the unknown parameters that need to infer.

3.2. 3-level CRP based Gibbs Sampling Inference Algorithm

Inference algorithms include two main categories: variational inference and Gibbs sampling. We prefer the latter for its fast convergence. Fig 3 shows the main phases in the Gibbs sampling. The main body of the algorithm is the iterative three steps: decrement update parameters, sample parse tree and increment update parameters. Updating parameters is relative simple, for the conjugate relationship between prior distribution and likelihood distribution. Only the idea of sampling parse tree is introduced here for simplicity. CRP is a classic process of inference by simulating an infinite number of tables in a Chinese restaurant which can seats an infinite number of customers. Due to the hierarchical structure of HDP-HMM-SCFG, we propose a 3-level CRP to sample parse trees for trajectories.

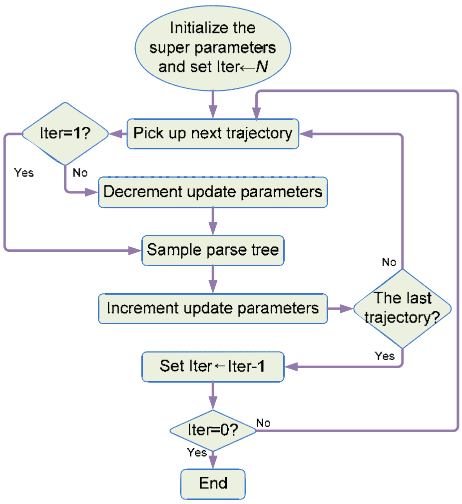

Fig. 3. Flow chart of the Gibbs sampling inference algorithm

On the top level, level of nonterminals, each nonterminal corresponds to a table; each pair $\varepsilon \in (N \cup T) \times (N \cup T)$ corresponds to the appetite of a customer; each multinomial distribution over the pairs, denoted by θ, corresponds to a kind of dish; the conjugate prior distribution $H = Dir(\alpha, \theta_0)$ corresponds to the menu. When a customer comes in the Chinese restaurant, he sits and chooses dishes under with the following three rules:

RULE1: The more customers that a table has seated, the more likely the table is chosen;

RULE2: The more extent that the appetite of the customers at a table match with the new customer, the more likely the table is chosen;

RULE3: One dish is chosen for each table, by taking account the appetite of all the customers at the table, until no more customers comes.

The rules can be formalized as equation (1).

$$P(z_{k+1} \mid \varepsilon_{1:k+1}, z_{1:k}) = \frac{\sum_{i=1}^{k} [P(\varepsilon_{k+1} \mid \{\varepsilon_l : z_l = z_{k+1}\}) \delta_{z_i}(z_{k+1})]}{k+\alpha} + \frac{\alpha P(\varepsilon_{k+1} \mid H)}{k+\alpha} \tag{1}$$

Where the z_i denotes the index of the table chosen by i-th customer, the marginal probability $P(\varepsilon_{k+1} \mid H) = \int_\theta P(\varepsilon_{k+1} \mid \theta) H(\theta)$ denotes the probability of a customer choosing a table, given the menu on the table. The conditional probability $P(\varepsilon_{k+1} \mid \{\varepsilon_l : z_l = z_{k+1}\})$ measures the extent that the k-th customer's appetite matches with others sits at the z_{k+1}-th table.

The situation on the middle level of terminals and bottom level of HMM components is similar, thus it will not be detailed redundantly.

3.3. Generative Classification

The likelihood of generating a sentence, given a SCFG, can be calculated during parsing. Similarly, the likelihood of generating a trajectory can also be calculated, given a HMM-SCFG. In the framework of grammar based classification, we learn a HMM-SCFG for each class of trajectories in train set independently. Then, the label of the class with the maximum likelihood to generate the test trajectory is assigned to it. See equation (2).

$$Lable(x_1, x_2, ..., x_l) = \arg\max_i P(x_1, x_2, ..., x_l \mid HMM\_SCFG_i) \tag{2}$$

4. Experiment

In order to test the effectiveness of the representation model, experiments are carried on the dataset of Austrian sign language trajectory (ASL). The ASL dataset is obtained from the University of California at Irvine's KDD archives [15]. There are 98 categories of 6757 trajectories acted by five signers. The dataset is chosen for two reasons. The first one is the challenge caused by the dataset's unusual complexity. The second one is that the same dataset as previous works will facilitate the result comparison. A half of the dataset are used for training SCFGs, the other half are used for testing the accuracy of classification.

Table 3. Accuracy comparison

Different Approaches	Number of classes									
	2	4	8	16	29	32	38	50	70	98
GMM [6]	0.98	0.89	0.85	0.74	0.67	—	0.64	—	—	—
HMM [6]	0.96	0.92	0.86	0.78	0.69	—	0.66	—	—	—
HDP-HMM-SCFG	0.97	0.94	0.86	0.75	0.61	0.58	0.49	0.37	0.29	0.13

Similarly with the work of Bashir et al. [6], we make classification on different scales and count the true positives to obtain the accuracy. The accuracy is the number of true positive divided by the number of all data. The data in Table 3 shows obviously that our approach outperform the previous work.

5. Conclusion and Future Work

We propose a novel model of HDP-HMM-SCFG for representing and classifying trajectories, and the result of the experiment on simulation data and ASL data has prove the effectiveness of our approach. The main contribution includes 3 main aspects.

Firstly, it's the first time that transplants the stochastic grammar based model to the domain of trajectory representation and classification. Compared with the traditional data driven models in domain of machine learning and pattern matching, our model is more capable of capturing the trajectories' hierarchical structure and temporal information.

Secondly, a 3-level CRP based Gibbs sampling inference algorithm is proposed to learn SCFG parameters. Compared with the traditional knowledge driven models, parameters of our model can be inferred from train set automatically, rather than predefine manually by domain experts.

Thirdly, the generative classification can not only classify a trajectory into some certain category, but also obtain the extent or confidence that a trajectory belonging to some category.

In the future, the grammar parse algorithm can be optimized to improve the efficiency of parsing. The methods for predicting future path given partial trajectory will also be further researched.

References

[1] Hu WM, Xiao XJ, et al. A system for learning statistical motion patterns. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2006, vol. 82, no. 9, p. 1450-1464,

[2] Hu WM, Xiao XJ, et al. Traffic accident prediction using 3-D model-based vehicle tracking. *IEEE Transaction on Vehicular Technology*, 2004, vol. 53, no. 3, p. 677-694.

[3] Hu WM, Tan TN. A survey on visual surveillance of object motion and behaviors. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications, and Reviews*, 2004, vol. 34, no. 3, p. 334-352.

[4] Zhong H, Shi JB, Visontai M. Detecting unusual activity in video. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, p. 819-826.

[5] Jiao L, Wu G, et al. The anatomy of a multi-camera video surveillance system. *ACM Multimedia System Journal Special Issue*, 2004, vol. 10.

[6] Bashir FI, Khokhar AA, Schonfeld D. Object trajectory-based activity classification and recognition using hidden markov models. *IEEE Transactions on Image Processing*, 2007, vol. 16, no. 7, p. 1912-1919,.

[7] Lawrence R, Hwang JB, *Fundamentals of Speech Recognition*. Prentice-Hall, New Jersey, 1993, SIBN 0-13-015157-2

[8] Vlachos M, Kollios G, et al. Discovering similar multidimensional trajectories. *Proceedings of the 18th International conference on Data Engeneering*, February 2002, p. 673-684.

[9] Buzan D, Sclaroff S, et al. Extraction and clustering of motion trajectories in video. *Proceedings of the 17th Conference on Pattern Recognition*, August 2004, vol. 2, p. 521-524.

[10] Warren LT. Clustering of time series data – a survey. *Pattern Recognition*, 2005, vol. 38, no. 11, p. 1857-1874.

[11] Berkhin P. A survey of clustering data mining techniques," *Technical Report of Grouping Multidimensional Data*, 2006, p. 25-71.

[12] Jelinek F, *Statistical Methods for Speech Recognition*. Cambridge, MA: MIT Press, 1997.

[13] Liang P, Jordan MI, Klein D. Probabilistic grammar and hierarchical dirichlet processes. *The Handbook of Applied Bayesian Analysis*, In pressed.

[14] Sethuraman J. A constructive de_nition of Dirichlet priors. *Statistica Sinica*, 1994, vol. 4, p. 639-650.

[15] S. Hettich and S. D. Bay, The UCI KDD archive. Irvine, CA: Dept. Inform. Comput. Sci., Univ. California.